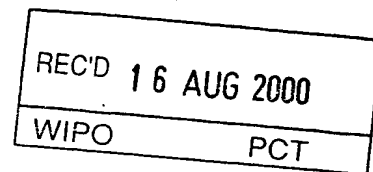


DE 00/1944

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)



4



**Prioritätsbescheinigung über die Einreichung
einer Gebrauchsmusteranmeldung**

Aktenzeichen: 200 07 494.6

Anmeldetag: 26. April 2000

Anmelder/Inhaber: Thomas Roitsch, Regensburg/DE

Bezeichnung: Promotorsystem, dessen Herstellung und
Verwendung

IPC: C 07 H, C 12 N

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Gebrauchsmusteranmeldung.

München, den 18. Juli 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Faust



Promotorsystem, dessen Herstellung und Verwendung

Beschreibung

Die Erfindung bezieht sich auf die Genexpression und die Regulation von Genexpression in Pflanzen. Im Speziellen bezieht sich die Erfindung auf DNA-Promotor-Sequenzen, und auf Expressionskassetten, die in Pflanzen eingeführt werden können, um die Transkription einer benachbarten kodierenden Sequenz zeitlich und räumlich innerhalb der Pflanzen zu regulieren. Zusätzlich bezieht sich die Erfindung auf Expressionsvektoren, die solch eine Expressionskassette enthalten und die benutzt werden können, um Pflanzen zu transformieren.

Hintergrund zu der Erfindung

Ein Promotor ist eine DNA-Sequenz, die Expressionsort und -menge eines Genes beeinflusst oder bestimmt, und die Stellen für die Bindung der RNA-Polymerase zur Verfügung stellt. Die Position eines Promotors ist im Genom eines Organismus relativ zum Transkriptionsstartpunkt fixiert. RNA-Polymerase ist ein Enzym, das an den Promotor binden kann und die Transkription eines Genes vollzieht, das unter der Kontrolle dieses Promotors steht. Dabei entsteht die messenger-RNA, die wiederum zur Synthese des Proteins verwendet wird.

Promotoren sind in verschiedenen Organismen untersucht worden. Für bestimmte Spezies konnten konservierte DNA-Bereiche (sog. Consensus-Sequenzen) innerhalb von Promotoren gefunden werden, die mit verschiedenen Genen assoziiert sind. Von diesen Bereichen wird angenommen, daß sie in die Rolle, die der Promotor im Transkriptionsprozeß spielt, eingebunden sind.

Die Initiation des Transkriptionsprozesses in Pflanzen schließt eine Interaktion des Promotors mit der RNA-Polymerase II ein. Innerhalb von Pflanzenpromotoren wurden Consensus-Sequenzen oberhalb des 5'-Endes des Transkriptionsstartpunktes gefunden. Eine dieser Sequenzen ist etwa 7 Basenpaare lang und befindet sich etwa 20-30 Basenpaare oberhalb des Transkriptionsstartpunktes. Diese Sequenz ist als sog. TATA-box bekannt und es wird angenommen, daß sie eine Rolle bei der RNA-Polymerase Bindung spielt. Eine andere Sequenz mit einer Länge von ungefähr 9 Basenpaaren ist etwa 70-90 Basenpaare oberhalb des Transkriptionsstartpunktes zu finden. Diese Sequenz wird CAAT-box genannt und es wird angenommen, daß sie in der Regulation des Transkriptionslevels eine Rolle spielt. Es wurden

3

nöch andere Regionen oberhalb des Transkriptionsstartpunktes identifiziert, die die Häufigkeit der Transkriptionsinitiation in Eukaryonten beeinflussen. Diese DNA-Bereiche, die Enhancer genannt werden, beeinflussen die Aktivität von Promotoren in ihrer Nachbarschaft. Diese Sequenzen sind jedoch der Definition nach keine Promotoren, da ihre Position nicht fixiert sein muß.

Um ein fremdes Gen in einem Organismus, z.B. einer Pflanze, exprimieren zu können, muß die kodierende Sequenz dieses Genes unter die Kontrolle eines Promotors gestellt werden und in die Pflanze eingebracht werden. Zur Insertion des zu exprimierenden Genes in das Pflanzengenom wird die fremde DNA meistens in das Ti-Plasmid von *Agrobacterium tumefaciens* gebracht, und dieses wird dann verwendet um die Pflanzen zu transformieren. Eine zweite häufig verwendete Methode ist die direkte Transformation von DNA z.B. mit Hilfe der sogenannten "particle gun". In den meisten Fällen werden hierfür bisher aus Bakterien isolierte Promotoren oder Promotoren von Pflanzenviren verwendet, die zur Expression des fremden Genes in den Pflanzen führen. Diese Promotoren haben den Nachteil, daß sie artfremd sind und daher den Kontrollmechanismen innerhalb der Pflanzen nicht unterliegen.

Bei Verwendung eines pflanzlichen Promotors ist die Expression eines fremden Genes möglich, die somit auch den pflanzlichen Kontrollmechanismen unterliegt. Durch Untersuchungen der Expression des Genes, vor dem der Promotor ursprünglich liegt, lassen sich genaue Kenntnisse über die Expressionsstärke, die Zeit, zu der das Gen exprimiert wird, und den Expressionort sammeln, die auf die Expression eines fremden Genes, das unter die Kontrolle dieses Promotors gestellt wird, weitgehend übertragbar sind. Ein weiterer Vorteil ist, daß bei Verwendung eines genau charakterisierten, pflanzlichen Promotors gezielte Eingriffe und Untersuchungen in die Entwicklung bestimmter Pflanzenteile möglich ist.

Im vorliegenden Fall wurde der Promotor einer Invertase aus Tabak kloniert, die in sehr großen Mengen, aber nur in einem bestimmten Entwicklungsstadium und hochspezifisch, nur in den Antheren exprimiert wird. Zur Klonierung des Promotors wurde eine genomische Bank aus Tabak mit einer Sonde aus dem kodierenden Bereich der untersuchten Invertase durchsucht. Die erhaltenen genomischen DNA-Sequenzen wurden mit molekularbiologischen Methoden weiter charakterisiert und schließlich ein Klon isoliert, der die Promotorsequenz enthielt, die dann sequenziert wurde. Es wurden Expressionskassetten mit verschiedenen Genen hergestellt, die in *Agrobacterium tumefaciens* eingebracht wurden, um Tabakpflanzen damit zu transformieren. Dieses neue Promotorsystem kann nun dazu verwendet werden, fremde Gene in Pflanzen in

4

einem bestimmten räumlichen und zeitlichen Rahmen zu exprimieren. Das Wort fremd meint dabei Gene, die nicht natürlich in Verbindung mit diesem Promotor vorkommen. Außerdem kann mit Hilfe dieses Promotorsystems die Expression von fremden und eigenen Genen in der Pflanze moduliert werden, d.h. Gene können überexprimiert oder reprimiert werden.

5

Ausführungsbeispiele:

1. Die Promotor-DNA-Sequenz, bestehend aus den 4300bp des 5' liegenden DNA-Bereiches, gezählt oberhalb vom Translationsstartpunkt der antherenspezifisch exprimierten Invertase aus Tabak, oder Teile davon, die dadurch gekennzeichnet sind, antherenspezifische Expression für dahinter gelegene Gene zu vermitteln, kann mit Hilfe molekularbiologischer Methoden so verändert werden, daß Restriktionsschnittstellen am 5'- oder 3'-Ende eingefügt werden.
2. Die wie in 1. veränderte Promotor-DNA-Sequenz oder Teile der Promotor-DNA-Sequenz können zur Herstellung einer Expressionskassette zur Expression fremder Gene in Pflanzen verwendet werden. So eine Expressionskassette ist dadurch gekennzeichnet, daß sie a) die Promotor DNA oder Teile davon aus 1. enthält; b) eine Verbindungs-DNA ohne spezielle Funktion bzw. fremde Gene, verbunden mit der ersten Schnittstelle beinhaltet; und c) eine 3'-Region, bestehend aus der 3'-Region eines eukaryontischen Genes enthält, wobei diese 3'-Region eine zweite Restriktionsschnittstelle an ihrem 5'-Ende besitzt und diese 3'-Region über diese zweite Restriktionsschnittstelle mit der Verbindungs-DNA bzw. den fremden Genen aus b) verbunden ist.
3. Eine Expressionskassette wie in 2. beschrieben, kann in einen Expressionsvektor kloniert werden. Dieser Expressionsvektor kann dazu verwendet werden, mit Hilfe verschiedener gängiger Methoden (z.B. Agrobakterien vermittelte Transformation; direkte Transformation) Pflanzen zu transformieren. Transgene Pflanzen können dann unter Bedingungen angezogen werden, unter denen das fremde Gen unter der transkriptionellen Kontrolle der beschriebenen Promotor-DNA-Sequenz exprimiert wird.
4. Die Promotor-DNA oder Teile davon können dazu verwendet werden, die Translation eines pflanzeigenen Genes zu modulieren. So kann die Expression eines Genes durch Einbringen weiterer Kopien unter der Kontrolle dieses Promotors gesteigert werden oder die Expression kann mit Hilfe der Antisense-Technik unterdrückt werden. Dazu muß die DNA-Sequenz des zu unterdrückenden Genes in "verkehrter Richtung" in eine Expressionskassette wie unter 2. beschrieben kloniert werden und wie unter 3. beschrieben in Pflanzen transformiert werden.
5. Die spezifischen Eigenschaften der Promotor-DNA-Sequenz bzw. von Teilen davon, ermöglichen in transgenen Pflanzen, die wie unter 3. beschrieben hergestellt wurden, eine

6

zeitlich (nur während der Pollenbildung) und räumlich (nur in Antheren) definierte Expression von fremden Genen in Pflanzen.

6. Aufgrund des starken Expressionslevels der antherenspezifischen Invertase in Tabak, lassen sich mit Hilfe dieser Promotor-DNA-Sequenz bzw. mit Teilen davon, in transgenen Pflanzen große Mengen eines bestimmten Proteins zu einem bestimmten Zeitpunkt und in einem bestimmten Ort der Pflanze (siehe 5.) herstellen. Dieses Protein kann dann durch Ernte der Antheren, Aufschluß und für das hergestellte Protein spezifische Reinigungsverfahren in großen Mengen gewonnen werden.

7. Die Promotor-DNA-Sequenz, oder Teile davon, können dazu verwendet werden, in die Entwicklung der Antheren in Pflanzen einzugreifen. So können transgene Pflanzen hergestellt werden, bei denen beispielsweise durch Antisense-Expression von Invertasesequenzen die Proteinmengen für die extrazelluläre Invertase verringert werden. Dies führt zu männlich sterilen Pflanzen, die in der Landwirtschaft, bei der Herstellung von Hybridsaatgut von großer Bedeutung sind.

8. Die Verwendung des Promotorsystems zur Herstellung männlich-steriler Pflanzen, wie z.B. unter 7. beschrieben kann als Sicherheitssystem bei der Herstellung anderer, kommerziell oder wissenschaftlich nutzbarer transgener Pflanzen verwendet werden. So besteht bei Verwendung männlich steriler Pflanzen nicht die Gefahr des Auskreuzens der genetischen Veränderungen auf Pflanzen die auf benachbarten Feldern oder wild wachsen. Die begrenzte Natur des Eingriffes der zur männlichen Sterilität bei Verwendung dieses Promotorsystems führt stellt einen besonderen Vorteil dar, da nicht in das vegetative Wachstum der Pflanze eingegriffen wird und keine Wechselwirkungen mit den zusätzlich eingebrachten genetischen Veränderungen zu erwarten sind.

9. Die Promotor-DNA-Sequenz, oder Teile davon, können dazu verwendet werden, transgene Pflanzen herzustellen, die pflanzeigene Stoffe in großer Menge herstellen, die positiv auf ihre Entwicklung, insbesondere betreffend den Ertrag von fruchttragenden Pflanzen, wirken können. Beispiele für solche pflanzeigene Stoffe wären Wachstumshormone oder Proteine die zur Energieversorgung der wachsenden Gewebe (z.B. Invertasen, Zuckertransporter) notwendig sind.

7

10. Mit Hilfe des Promotorsystems können transgene Pflanzen hergestellt werden, in denen die Menge der produzierten pflanzeigene Stoffe (z.B. Wachstumshormone) reduziert werden kann. Diese Reduktion kann durch Einbringen von abbauenden Enzymen, von Inhibitoren oder durch sog. "single-chain" Antikörper erreicht werden.

11. Um Erträge von männlich sterilen, fruchttragenden Pflanzen zu erhalten und zur Vermehrung männlich steriler Pflanzen, die unter Verwendung des Promotorsystems wie z.B. unter 7. beschrieben hergestellt wurden können Restorerstämme hergestellt werden, die nach einer Kreuzung mit den männlich sterilen Stämmen zu fertilen Pflanzen in der F1-Generation führen. Restorerstämme für die unter 7. beschriebenen männlich sterilen Pflanzen könnten Proteine enthalten, die die Kohlenhydratversorgung der Antheren wieder herstellen können, wie z.B. artfremde Invertasen (z.B. aus *Saccharomyces cerevisiae* oder Bakterien), oder Saccharosetransporter in Verbindung mit intrazellulären saccharospaltenden Enzymen (z.B. Saccharose Synthase, neutrale oder vakuoläre Invertasen).

12. Alternativ zur Herstellung von Restorerstämmen wie unter 11. könnten Pollen von männlich sterilen Pflanzen in einer in vitro Kultivierung zu fertilen Pollen entwickelt werden, mit denen dann eine Befruchtung der transgenen Pflanzen stattfinden kann.

13. Pollen können durch eine in vitro Embryogenese zu haploiden Pflanzen herangezogen werden. Diese haploiden Pflanzen können dann zu homozygoten diploiden Pflanzen gezüchtet werden. Zur Induktion dieser in vitro Embryogenese ist ein Hunger- und Stress-Schritt nötig. Da transgene Pflanzen, die unter Verwendung des Promotorsystems wie z.B. unter 7. beschrieben hergestellt wurden, in ihrer Zuckerversorgung gestört sind, sind die Pollen dieser Pflanzen bereits in Richtung Embryogenese determiniert und es bedarf keiner zusätzlichen Hunger- oder Stress Behandlung mehr. Die Embryogenese läuft in diesen Pflanzen daher schneller und effizienter ab.

Erreichte Vorteile:

Mit Verwendung dieses Promotorsystems steht ein Werkzeug zur Verfügung, mit dem man zeitlich und örtlich gezielt fremde Proteine in Pflanzen exprimieren kann und pflanzeigene Gene in ihrem Expressionslevel modulieren kann.

Ansprüche

1. Promotorsystem, gekennzeichnet durch einen Klon aus genomischen DNA-Sequenzen.
2. Promotorsystem, gekennzeichnet durch Expressionskassetten.
3. Verwendung eines Promotorsystems nach Anspruch 1 oder 2 zum Beeinflussen eines Genes.
4. Verfahren zur Herstellung eines Promotorsystems nach Anspruch 1 oder 2.
5. Verfahren zum Experimentieren mit einem Promotorsystem nach Anspruch 1 oder 2.

Zeichnung 1: Promotor-DNA-Sequenz der extrazellulären Invertase aus Tabak

```

1   TCTAGAAATGA CGCCACCGGC CAGGACGGGG AGTATGATTT CCCCGAATGT
51  TCGTTCAACT GCATTGTAA AACCTGTTAG CGTGATGCAG CCCGGTACTA
101 TCTTATCCTC GAGTTCATT TGTGCAAGTA CTCGAGGATG GACAATTCAC
151 GGGCCACTCC CATCGTCCAC CATAATGCGT CTTACATCTG TATCTAATAT
201 TCGTAAAGTG ATAACGAGGG CATCATAGTG AGGGAAAACC AAACCGTGGT
251 TATCTGACTT ATCGAAGATG ATACTTTCTT TAAGTTTCTC GTACCGTTCFA
301 TGAGTGATTA ACTGTTTGAG CTTGTGGGTT GTGGCGAAGT TTACGTTGTT
351 GATCGAAACG TCGTCTCCGC CCCCGATGAT AATGTGAATG GTGCGAGTCG
401 GTAAGGGTGG TTTCCGGCGGT CCTCGGTGTT GTTCACGTCC TCGAGAAAAG
451 TTGGTCCCTC CTCGGTCACA CAACAATATT TTGAGGTGTC CTTGATGAAG
501 CATGTCCATG ACCCTCTGTC TTAGGGCGAT ACAATCCTCA GTTTTGTGAC
551 CTCGCTCTTG GTGGAAGTCG CAGAGGGCAT CTGATTTTCT AGTGCTTGGA
601 TCTGACCTCA TCTTTTGTGG CCACTTTACT TTTGGTCCGA GCTTCTTCAA
651 TGCATAGACT ATTTCTGAGG GTGACACACA AAATTTGTGA GCGGATAGTA
701 AAGAGGGCAT ACCTCTCTCG TTCCGGTGAG TCCCTGTCTT TGGCCTAGAT
751 GGGCCCTCTT CGTAGCGGGA GAGGGGCATG ATGGCACTTT TGACATATGG
801 TTGATCCATT TCTCGGTTAG ATCATGGAGC TGCAAGATCT CTCTTGGCAT
851 CATTTTGACG ATCCTTCCTG GTTTCGGCTT GTACCGAGGT CAATCGATGA
901 GTTGGCCGAT TCAGGTCGTC TTCGTCGGCA CGGGCCTCAG CACAGTAGGC
951 GTTGTGTATT TCATCCCAAG TGGTTGGAGG ATATTTCTAT AGTTGGTTTA
1001 ACAGTTTCTT GGTCCGCCCT GAGCCATTCA TGTTGAGCCC ATTCTGGAAA
1051 GTTGCTACAA CCATTCCTTC TGATACATTC GGTAAGGTCA TCCTTACTCT
1101 GTTGAATCGA GCGAGGAAGT CCCTCAATCC CTCTCCGAGT GATTGTTTGA
1151 TGGCAAATAT ATCGTTCACT CTTGCCTCCG CGTTTTTAGC CCCAACATGG
1201 GCCATTATGA ACTTGTGCGC CATCTCTTCG AATATTTCAA TGGAGCGCGC
1251 GGGCAGCTGT GAATACCAAG TCAATGCTCC TCCGGTAAGG GTCTCGCCGA
1301 ACATTTTCAA CAAGATGGAG GAGACTTGTT CTTTGGAGAG ATCATTGCCC
1351 TTACCGCAG TGACATAATG ATTACATGAT CTTCGGGGTC GGTCTGACCA
1401 TCATAAATTT TCAGATAAGG TGGCATCTTG AACGTCTTGG GTATGGCATA
1451 TGGGGCGGCT TCATCACTGT AGGGTTGCTC GACTAACCGA CCAGCGTCTC
1501 TTTTGGAAA TATTTTGGG GCACCCGGTA TTTTATCGAC TCTTCTTGG
1551 TGTTCTCTCA TTTGATCCCG AAGCATTTTA TTTTCGTTTT CCATTTCTTC
1601 CATTTTCTTC AGAATGCCCG TGAGGGTGTG ATTACCTGCA TTATTAATAT
1651 TGTGAGTGAT ACCTGTTACT GAAAGGGGAG GGTCTGTCTG TTTGGTCAAT
1701 GCTGGTGCAA TGCAAGTCTT TGCATTTTCT CTAATAACCT CCTGAGTGGG
1751 TTTGTTGAGG ATGCCGGTCA GCATATTTGT CAGCCAAGCT TCGAGTAGCT
1801 TCTTCACCGC TGGTGGCGCC TCTTCCGTTG TGGACGTGGA AGCTCCTTTA
1851 CCGCGGGATG TTGCGATACT GCTGTGAGGG AGGGGTGATC CACTTCGTCT
1901 GGGAGAGGTG TTAGGCGTTA TGCCCTCGCC TTCTATTTCT GAGACCTCAT
1951 TGATGGTGTG TAAGAGGTTG GTAGTGAGAT TGGCCACTGC CTTCTATCCTT
2001 TCTTCTCCCT TACCTGCCAT GTCAGATCTG GGTGTACAAG GAAGTAGGAG
2051 CTTCTCTTCT TCTTTTGTGT GAATTGTGCC AGTTATAGAT CTAAAAGAAA
2101 CTAAAGTTTT AACTAGACTA TCCTCACAGA CGGCGCCAAA TTGTTTGACC
2151 AAAAAATATA GACTTTTGAT TAAATTAATT AATATTGTAT GACAAAGGAT
2201 TAAACCTAGT TAATGATAAT AACTTCAGAT CTATAATCAA TTAACAGCAA
2251 TCACGGTCAT AGCAGCGTTG AGAGAAGATT AAATGTGATG TYCATTCAT
2301 ATTTCAAGAT CATTAATGAT AGGGGAATAT CAAGCAATAA ATAACGATAA
2351 ATGGCATTAAG AGTAAATAG GAGAATGATT CACCCAATAT TGAATGAGGT
2401 GGATGATTCT TCTTTTGTAC AATGATGAAT GATGGGCAA TACTAGAATG
2451 TTGGGACCCT TCTCGGATCT AATGAAAAAA GTATGGAATA GTAGATAATC
2501 GAATCTCTTT AGAAAGGTAG TGATTGTCTT TTATCTAGAG AGAAAGTCTG

```

10

2551 CTTTTCAAAG AATATTTTTA TCAGAGAATA TTACATCCCC CTCTCTCCCT
2601 ATCTCTTTTT CTATTATAT GGGACATTC TGAATCAATC CTAAAAGTAC
2651 ATACACCAAG AATATTCAAT AAAATATTT TTTGAATATT CTATTATAAA
2701 AACTAGCTGT TAGCACTCGA CCTGGGTCGY TATTGACTAC TCGGTTACGA
2751 GCCCTGTCAT TTACTAATCG ACCTCGATTA CATCACTTTC TACGATACTG
2801 CTTTCATGTCA AATCTTAATG AAAGCAGATT TTGACCCATA CAATAATATG
2851 ACAAATGTC TTCCAAAGAA AACATGGCTC TTATAGTGAA ATATCGTTAG
2901 ACTGTTATAG AAAGATCTGA ATTTATTTAT AAGAATAGTG TTTTTTCTT
2951 TTCTTTTCAT ATCTAAGGAG TAAAGCAACC ATGAATAGAA AAGGCTTAGT
3001 AACTATATAT CAAAGGAATG GTGTTTTTTC TTTAAATATG GATAAAAATT
3051 TGTGAATATA GAAGATTAGA TCAATTAACA AAGGTTATGG TGGAGTGGTA
3101 AGCAGAGGCG GACCTATGTG TTATAGTAAG GGGTCACCCA CTACTAGAAA
3151 TCCGCTAAAG ATCGATCAAA AAACCGACCA ACATTGGTCG GTAATGGCCA
3201 AAAACTGACC AAACCGGAT CATTTACGTG TGAACGGTAT TTTTATGGTC
3251 GGAAAGGAAT ACCGACCAA GTTGGTCGGA AATTACCGAC CAACTTTGGT
3301 CGGTCAATTA AATTCAAAAA AATATTGTA AAAAAAACC GACCAAAGTT
3351 GATCGGTATT TTAATTATGT AATAAAAAGA TTCACTATCT GGGAATCGAA
3401 CCGGGTCTG TACTATGGCA AGATACTATT CTACCACTAG ACCATTGGTT
3451 CATTTTGTGTT TAAGACTGTC TTTTATTTGA TTTATACTCT TTAATTATAT
3501 TTTTGCACGA AATAACCGA CCAAAGTTGG TGGATTTTAT TAAAAAGTAA
3551 AATTACTTAC CAAAGTTGGT CGATTTTTTT AATGATCCG CCGAATTAAC
3601 CGACCAATTT TGGTAGGTTT TTTAATATT AATTTTTATT TATTTTAATT
3651 GAAAAACTAA CCAAAGTTAG TCGCTTTCTT GAAACATAAA TTTCGCGGGA
3701 CTCAAAAAATA GTTCCCGCA TTTTTCGCC AAAGAAAACC GACCAAAGTT
3751 GGTGCGTTTC GTAAAAAAA AAAAAATTTA AAAAAATAT TTTAAAAAAC
3801 CGACCAACTT TAGTCGGTTT TTTGGTCGAT TTTTTCACCG ACCAAAGTTG
3851 GTCGGTCGAC CTTGGTCGGT TTTTCCCGAA TTTGTAGTAG TGACCGAACC
3901 CTGTAAGCTT CCGGAGAAAT TTTGTATATG TATATGTGTA TATCCTTAAA
3951 ATGATTAAT TAAAGAACGT GGCACCCTGA ATACTAGAAG CCTTLAGGGG
4001 CACTAGATGA GCAGAATAAC GTGTTCTCGT CGCGTAAAAA TACTTGGATC
4051 CGCCTATGAT GGTAAGTACT TCTTCGTCTT TAATCAGAGG TTTTCGACTTC
4101 GAGCTCCAGA TATAACTAT AGACTCGTCT TTATAGCACC TTTTAATAAG
4151 ACTATGACT CATCTGATTT CTCTATAAAT ACTCCTCAAG CTTTCGGTTC
4201 TTCTCCATTG TTCAGTTTCT TCTCCACAT CACAGAAGTG AAAACAAAAC
4251 AAGAAGAAGA AGAAGAAGAA AAATAAAGAG TTTCTGTCAA ATTAAGTCCA
4301 ATAGGGAAAA TG